AD_____

Award Number DAMD17-94-J-4414

TITLE: cDNA Fingerprinting of Breast Cancer Tumor Cells

PRINCIPAL INVESTIGATOR: Cassandra L. Smith, Ph.D.

CONTRACTING ORGANIZATION: Boston University
Boston, Massachusetts 02215

REPORT DATE: October 1998

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1998	Final (30 Sep 94 - 30 Sep 98)

4. TITLE AND SUBTITLE
cDNA Fingerprinting of Breast Cancer Tumor Cells

5. FUNDING NUMBERS
DAMD17-94-J-4414

6. AUTHOR(S)
Cassandra L. Smith, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Boston University
Boston, Massachusetts 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

19990405 031

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research is developing and applying comparative DNA analysis to breast cancer disease. For instance, methods for making DNA chip arrays useful for matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) analysis were developed. Targeted genomic and cDNA differential display (TGDD and TcDD, respectively) will be used to feed sample hungry MS instruments pools of fragments containing a common sequence (e. g. $(CAG)_n$, or other repeat sequences, LTR sequences (retroviral footprints), or Zn-finger binding motifs (transcription factor coding sequences). Targetings is on important dispersed gene and sequence families whose expresson, or genomic organization may be modified in tumor cells. Other array oriented technology developed include rolling circle amplification (RCR: an *in situ* isothermal DNA amplification) and an *in situ* scoring method for genetic markers. Cloneless libraries generated from restriction enzyme cleaved genomic DNA fractionated by electrophoresis were used to characterize the chromosome 20q13 region amplified in breast and ovarian cells, and to identify ~70 cDNAs from this region. Gel lane slices containing DNA constitute the library.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

__✓__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

__✓__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

__✓__ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

__✓__ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

__✓__ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature Date

Table of Contents

Introduction

The long term goal of these experiments is to develop and apply methods that can identify DNA sequences that are important for diagnosis and treatment of breast cancer. The methods developed by this work include: (1) mass spectrometric analysis of DNA arrays, (2) isothermal rolling circle amplificaton, (3) solid state scoring of simple repeat sequences, and (4) genomic analysis with cloneless libraries. Method (1) was tested with on tumor and normal cells from the same individual. Method (4) was applied to analyzing chromosome 20q13 region amplified in some breast and ovarian cancers. All methods are still under development but significant progress has been made in understanding and solving the problems that have plagued other researcher who have attempted similar experiments. The development of truly novel methods that are robust over a wide range of conditions always takes longer than expected.

Genetic studies on the heritability of cancer seek to identify causative DNA sequences. The first stage of positional cloning experiments use genetic mapping to identify a, hopefully small, region within which lies the sought gene(s). Then physical (molecular) methods are applied to further narrow the search region, to identify genes within the search region and finally to identify the disease gene. Recently, these methods were used to isolate the BCRA1 (Harshman *et al.*, 1995) and BCRA2 (Tavtigian *et al.*, 1996) genes. It is estimated that over ten years ago it costs about $120 million dollars to find the single gene involved in cystic fibroses. Since then, The Human Genome Project has provided an increasing number of resources for finding disease genes using positional cloning methods. Still, the task of finding genes involved in diseases is arduous and expensive. Most recently, Myriand Genetics estimated it cost $24 million by that company alone to find the first familial breast cancer gene.

Thus far, the positional genetic approaches have only identified major gene causes. Whereas the onset or progression of many diseases is governed by multigenic effects and interactions. Even major disease genes are not expressed alone but in a chorus of over 80,000 other genes. Given the spectrum of genomic changes thus far indentified in breast and other cancers, it is quite clear that more efficient methods are needed to

analyze the increasing number of genomic sequences important in tumor development, progression and response to therapeutic regimes. Thus, a number of groups, including ours, focused on developing comparative methods for identifying multi-gene differences between samples that can be applied in a cost effective method to a large number of samples and on increasing the efficiency of positional cloning experiments.

Although the published methods for multigene analysis are useful as research tools, none have proven to be robust enough to be routinely applied to samples that have the complexity of the human genome. The approaches include comparative genome hybridization (CGH; Kallioniemi *et al.*, 1994), differential display (DD: Liang and Pardee, 1992; Liang *et al.*, 1994) and subtractive hybridization (Lisitzyn *et al.*, 1993a; Lisitzyn *et al.*, 1993b). In CGH, an equal molar end concentration of differentially labeled cDNAs from two samples are simultaneously hybridized to metaphase chromosomes. Genomic regions that are amplified or deleted in one of the test samples will be differentially labeled. Hence, this method can identify genomic regions important in disease states.

In DD experiments, mRNA levels of appropriate samples are analyzed. Here, total mRNA from different samples is amplified randomly and displayed by size, electrophoretically. The differentially expressed cDNAs are then isolated and characterized. In subtractive hybridization, sequences present in one cDNA library but missing in a second cDNA are isolated and characterized.

An alternative method of measuring mRNA level is the random sequencing of cDNA libraries made from particular cells. Although several pharmaceutical groups with large resources are taking this approach for some diseases, it is quite clear that DNA sequencing costs at this time preclude the use of this method for routine application. Our original proposal intended to extend the principles of CGH to arrays of cDNAs. Since this proposal was written two methods for differential display of cDNA were described. One method (Schena *et al.*, 1995) is very similar to that described in our original research proposal. The method involves hybridization of differentially labeled cDNA simultaneously to the same array of cDNA probe samples. Schena *et al.* (1995) reported on the application of CGH principles to arrays of yeast cDNAs. We also carried

out a number of pilot studies on several arrays of cDNA. The other method (Velculescu *et al.*, 1995) to quantitate gene expression uses direct DNA sequencing of chimeric small clones that are composed of ligated pieces of cDNAs. Each of the ligated pieces is an index for a particular cDNA. Thus, one sequencing reaction gives information about many cDNAs. The chimeric clones are created in a manner that should preserve quantitative information on the occurrence of each cDNA.

Body

Novel robust methods like those described here are difficult to develop. However, enormous progess has been made in identify obstacles and successfully designing methods around them. Although, some obstacles remain (especially for the implementation of the array technology) some of the developed methods were used to analyze the 20q13 region which is amplified in some breast and ovarian cancers.

(1) Mass spectrometric (MS) analysis of DNA arrays:

Our progress on DNA arrays can be divided into aspects (a) targeted genomic and cDNA differential display (TGDD and TcDD, repectively) and (b) mass spectrometry. The development of MS for DNA analysis involves the expertise of collaborators for instumentation, MS, chemistry, molecular modeling, engineering, biochemistry, biology etc. The DOA grant monies only pays a portion of the total cost of this program spread over several universities and industry. Our contribution to this collaboration has been developing methods to provide informative samples for analysis.

(a) TDGG and TcDD: TGDD and TcDD focuses analysis and reduces sample complexity by capturing genome subsets (i. e. restriction fragments) that contain a targeted interspersed repeat. Two methods have been discribed in method I (Broude *et al.*, 1997). Fragments containing the target sequences are captured by hybridization to an immobilized complementary single strand probe sequence. The captured fragments are labeled with fluorescein and amplified by PCR, and then fractionated by size on an automated DNA sequencing instrument. A second method was developed that is based solely on PCR (Method II, Broude *et al.*, 1998;

Oliverai *et al.*, 1998; Nguyen *et al.*, 1999). Method I and Method II produce different types of fragments for analysis. Method I produces fragments which contain the target sequence surrounded by unique sequences. Method II produces fragments containing the target sequence at one end of the fragment. Thus far, the sample pools have been analyzed electrophretically. Hence, the DNA fingerprint consisting of a display of the size distriubtion of restriction fragments containing a common target sequence.

Conventional DD analyzed cDNAs. The focus on cDNA provides a sample complexity reduction and a focus on interesting genomic subsets. The problem with cDNA analysis is that the dynamic range of expression is 10^5. It is difficult to maintain quantitative information when comparing samples of this dynamic range, especially when an an exponential amplification system (PCR) is used to generate the samples that will be analyzed. In DD, random cDNA are amplified and labeled by PCR and analysis by high resolution electrophoreses. This means that when mRNA is studied the sampling will only be on highly expressed genes. DD has been called "differential dismay" because of the high number of false positives (Debouck, 1995). Usually this problem is addressed by retesting individual differences before extensive characterization.

A goal of our experiments has been to minimize the number of false positives by identifying their causes. This allows us to obtain quantitative assessments of the difference between samples. Our experiments focused on genomic DNA instead of cDNAs. This means that the dynamic range of the sample concentrations being compared is very small (0 - 2) differences and that differences should seen in integral amounts. Many of our experiments were done with DNA isolated from monozygotic twins, or from different tissues from the same individual (mostly rat samples but some tumor vs tissue samples). Such samples should be identical or near to identical. Hence, unlike other similar studies which repeated experiments with the same sample, we focused on comparing different samples which should be identical or close to it. Our focus was on minimize differences between samples, so that when differences were detected that were likely to be real. Most recently, a model system using the *Saccharomyces cerevisiae* genome (Goffean *et al.*) was established (Bouchard *et al.*, 1999). Since the entire sequence of this

genomic DNA is known, experimental results can be compared to theoretical results. This allows an understanding of how incorrect results develop.

It is clear these type of experiments are difficult require close attention to detail. Several factors that are critical have been identified (Storm *et al.*, 1999; Nguyen *et al.*, 1999; Bouchard et al., 1999) . The results show that successful experiments can be carried out at high concentration of $MgCl_2$. The addition of high concentrations of $MgCl_2$ to each samples minimize differences due to in the amount of Mg^{2+} that are chelated to the DNA samples. Most importantly, it appears that the final DNA concentrations in the PCRs must be extremely closely matched. This is because the multiple sample peaks are not uniformly amplified during PCR (Nguyen *et al.*, 1999). This results is surprising and not clearly understood.

During the course of these experiments we developed quantitative computational methods of analysis (Bouchard *et al.*, 1999). These methods identified peaks and then calculated the area under each peak automatically. The quantitative methods were used to analyze results obtained with the *S. cerevisae*genomic DNA. The approach allow us to obtained quantitative results after several modifications to the procedures. The modifications included a change from the TAQ polymerase to the EXPAND PCR Enzyme Mixture obtained from Boehringer Mannheim. The TAQ polymerase has no 3' exonuclease activity (= proofreading activity). This means that the misincorporation of a wrong base prevented further extension of the template. The misincorporation rate for TAQ polymerase is rather high (1×10^{-5}). This can be overcome by the addition of a second polymerase (PWO) which has a 3' exonuclease activity to the highly processive and efficient TAQ polymerase as in the EXPAND system. Further, unwanted amplification products were eliminated when the PCR primer was modified to contained a sulfur substituted diester bond. The presence of the sulfur prevented the removal of the terminal base when it was mismatched.

A large number of our experiments were conducted on DNA isolated from blood samples of monozygotic twins. This provided a large amount of human samples that could be compared under a variety of conditions.

Furthermore, these could serve as a model for experiments comparing different tissues from the same individuals. In the later experiments smaller amounts of samples would be available, hence it was important to perform initial experiments on samples where a large amount of material was available. Several other samples were also examined. For instance, a series of experiments were done on different tissues isolated from the same rat. These experiment focused on studying the stability of the genome in general, in preparation for studying it in tumor cells. Another set of experiments were done using human lung and sarcoma samples. Here, once again the sample was choosen because a large amount of tumor sample could be obtained. A number of differences were documents in all of the comparative cases. We also attempted to analyze and compare some breast cancer tumor cells from paraffin embedded samples. The DNAs that were provided to us were too degraded to be useful. We are seeking higher quality samples. This may mean that we will need to improve the DNA extraction procedures used for paraffen embedded samples. Recently we have also made arrangement to obtain breast cancer biopsy material. This means that although the methodology could still use improvement, we now know enough to apply our method to breast cancer tumor cells.

Most of the experiments described above targeted $(CAG)_n$ or $(CA)_n$ which are known to be unstable in cancer cells. Also developed were targeting methods for LTR sequences which fingerprint the location of retroviral sequence and Zn-finger binding motif sequences. The method was also extended to include the analysis of cDNAs. The next target sequence will focus analysis on the signaling cascades that are so important in tumor biology. In particular we are currently developing our targeting protocol for classes of G-protein coupled receptors.

The methodology still needs improvement. For instance, we are still exploring the variables that affect the reproducibility of our genomic and cDNA differential display method. These are very tedious experiments that represent an enormous amount of work but absolutely necessary when robust methodology is developed. These experiments involve testing of all of the reaction components against each other in each of the steps to learn the optimum concentrations and incubation times and to learn the error bars allowable on each of variables. We are also continuing our development of methods for automatic computational methods of analysis

the similarities and differences in our display methods. This will allow us to evaluate different experimental approaches and to determine the level of differences between samples.

The long term objective of this research is to develop simple but accurate methodology that can be used to analyze large regions of the genome so that changes at the DNA or RNA levels associated with specific breast cancer characteristics can be uncovered. These changes may occur through point mutations, or larger DNA rearrangments on amplifications. Although a number of similar approaches have been developed and applied to clinical samples, most if not all of the approaches are either too expensive or too technically demanding to be of wide spread use. In contrast, our approach may be applied to a large number of samples. Discounting, salary we estimate a cost of about $20 per sample. A single techinician could hundreds of samples per month.

(b) MS Analysis of DNA Arrays: Only recently has MS analysis been applied to DNA (Graber *et all*, 1998). The masses of the bases are 289, 304, 313, and 329 for C, T, A and G, respectively. The accuracy of MS-TOF is 1 part in 10^3 (Note, that the accuracy of Ion-Cyclotron-Resonance (ICR)-MS is 1 part in 10^5 although this instrument is 10-fold more expensive). This type of accuracy allows the base composition of a DNA to be determined from its mass. An oligonucleotide of length L can have (L+3)!/L!3! different possible base compositions. Hence, array technology can be used to sort pools of DNA fragments and mark them with a known sequence index. Positional sequencing by hybridization (PSBH: Broude *et al.*, 1995) was developed by us to index the sequence at ends of fragments with great accuracy. The discrimination ratio between matched and mismatched sequences is not greater than 2-fold in convential array technology. In PSBH the discrimination ratio ranges up to 200-fold. Hence, capturing and indexing end sequences has proven to quite valuable for array technology.

Our approach to analyzing cDNAs by MALDI-TOF MS is to focus on specific gene classes provided by the methods described above. Hence, we will

adopt some indexing technique for sorting the generated targeted fragments to array elements for analysis. This combines known and unknown sequence elements in the analysis. A large number of groups are exploring indexing methods. Each method for preparation and selection has its own idiosyncrasies. However, the underlying steps are the same. Most work has been an expression profiling.

Generation of an expression profile involves the creation of cDNA samples using reverse transcriptase. Each cDNAs and genomic DNA requires a unique index of 10-15 nucleotides. PCR is carried out on all the targeted fragments in parallel and then the relative abundance of each indexed member is measured. It is clear that there are a number of way of indexes. Hence, some experiments have developed the necessary software tools, simulational and (data) analytical that are needed for modeling the various approaches.

Sample complexity reduction will be done through targeting and will be an intrinsic part of the indexing scheme. Array hybridization will be used to sort the targeting products to complementary array elements. This method combines some features of both indexing as originally suggested by Velculescu *et al.* (1995) with procedures used after more traditional rtPCR as described by Kato (1995, 1996) and Unrau and Deugau (1994). Each index fragment will be generated such that one (single indexing: SI) or both (double indexing:DI) ends have a single-stranded overhang. In each case, one end of the fragment will be hybridized to a spatially separated array of fixed hybridization probes; each probe has a unique single-stranded overhang, and each is analyzed separately by MS. The fixed probe array contains 4^m elements, where m is the number of nucleotides in single-stranded overhang. Our experiments (Broude *et al.*, 1994; Fu *et al.*, 1995) have shown that this greatly reduces the probability of mismatches between the anchored probes and their targets.

Further differentiation of DNAs is dependent upon whether SI or DI indexing is used. In SI, further differentiation is obtained through mass measurement. In this protocol, only one strand (length N) of the DNA is analyzed in the MS. Since, m nucleotides are known from the position in the array, this leaves $N-m = k$ nucleotides to be determined by MALDI-MS. In a DI approach, a mixture of specifically designed floating probes is

hybridized to the second single strand overhang after the cDN fragment has been hybridized into place in the array. For quantitative analysis, competitive hybridization can be used with a mass-labeled set of standards for each array element.

Simulation experiments will guide and optimize the accompanying experimental program which will be focused on examining the most serious error sources (1) accuracy of mass measurement by MALDI MS, (2) hybridization of slightly mismatched probes, (3) the quantitative represention of mRNAs by the RT-PCR generated cDNAs, and (4) the coincident occurrence of identical or nearly identical mass labels on different mRNA species.

These modeling experiments will also take advantage of the the National Cancer Institute's, Cancer Genome Anatomy Progect to include the ever increasing number of genes that have been identified to play some role in breast and other cancers. Eventually, these genes will make up another of our MS test systems since differential display has already been used to assess the level of these genes in about 20 different breast cancer cell lines and primary tumor cells.

(2) Solid state scoring of simple repeat sequences: Genetic mapping experiments require the analysis of an enormous number of genetic markers. Many of these markers are simple repeat sequences such as $(CA)_n$ or $(CAG)_n$. Repeat length is measured electrophoretically. The electrophoretical size fractionation is the rate limiting step. This step was replaced by an *in situ* scoring method (Yaar *et al.*, 1997; Surdi *et al.*, 1998).

The *in situ* scoring method uses immobilized probes. The probes are complementary to the target sequence. An array of probes has the same unique sequences but different length of simple repeat. A perfectly matched duplex is formed between probe and test DNA when the number of repeat sequences is equal. A mismatch duplex with a loop structure is formed when the probe and test DNAs have different repeat sequence length. The presence of the loop structure can be detected by S1 nuclease or T4 endonuclease VII which cleave the DNA at the mismatch. Single strand breaks introduced by the S1 nuclease can be nick translated to

remove or add labels to the duplex DNA. T4 endonuclease VII makes double strand breaks which can be used to remove or add a label. The advantage of using the T4 endonuclease VII is that a single enzyme is used. The disadvange is that there is a high signal to noise ratio and the enzyme is not very stable. The advantage of using the S1 system is the stability of the components and the very low background noise. The disadvantage is that in some implementation a second enzyme is need (DNA polymerase).

This appraoch is especially powerful when di- or tri-allelic system are characterized as is done in inbred ~~mo~~use mapping experiments. MS analysis of such samples may not even require enzymatic manipulation but simple hybridization. In this case the markers with specific types of simple repeat sequences could be captured and amplified using TGDD. The mass of the test sample would reveal the length of the repeat directly.

(3) Solid state isothermal rolling circle amplificaton: PCR technology has revolutionized molecular studies. The problem with PCR is that the products are soluble. This means that the products float away and positional information of an immobilized template is not retained. This means that PCR cannnot be applied to arrayed samples because of diffusion of the products. PCR requires cycling between at least two and usually three different temperatures. The high temperatures used in PCR destroy templates, enzymes and precursors. Lastly, the expotential nature of PCR that allows one to begin with very small amounts of templates also makes it difficult to retain quantitative information.

Rolling circle amplification (RCR) was developed by us to overcome many of the drawbacks of PCR (Hatch *et al.*, 1999). RCR is an isothermal amplification system that uses an immobilized primer. This means that the single stranded product is attached to the primer and that positional information is retained at the end of the reaction. This system was first developed on magnetic bead model system then transferred to silicon chips. The silicon chips were engineered to contain nanowells lined with streptavidin (Sabayana *et al.*, 1999). A 5' biotinylated primer bound by the streptavidin and extended by polymerase when a circular template was present. The test DNA was circularized by incubation of ligase with the immobilized primer and a single stranded target whose ends were complementary adjacent sequences on the primer sequence. It should also

be noted that this work demonstracted that macromolecular reactions could be done on silicon surfaces using immobilized substrates.

RCR allows the applification of target DNAs *in situ*. Furthermore, during RCR it is possible to add MS (or other) labels so that ligation and/or amplification is detected.

(4) Genomic analysis with cloneless libraries: Genes and other important sequences important in breast and ovarian cancer can be isolated from genomic regions identified in positional cloning experiments. Usually the first step in characterizing the region is the detailed molecular characterization of large insert clones from this region and the constrution of genomic restriction maps. Here, we have used genomic DNA directly in place of large insert clone libraries. This studied mapped genomic *Not* I restriction fragments on chromosome 20 and then focused on a region of human chromosome 20 amplified in breast and ovarian cancer tumor cells (e. g. region 20q13). This region was identified by CGH experiments of others (Tanner *et al.*, 1994) who have then used time consuming conventional positional cloning approaches to identify putative genes important in breast cancer (*Collins et al., 1998*).

Our approach uses pulsed field gel (PFG: Schwartz *et al.*, 1983; Schwartz and Cantor, 1994) fractionated genomic restriction fragments as a direct source of DNA (Mass *et al.*, 1999). Genomic DNA that cut with an infrequently cleaving restriction enzyme is fractionated by PFG under appropriate conditions (Smith *et al.*, 1992). The gel lane containing DNA is cut into 2 mm slices. Each slice is melted in a solution containing a preservative (20 mM of ethanolamine) by heating to 95 C for 15 min. These samples can be stored indefinitely. The DNA in agarose can be used as a template in a number of reactions including PCR. For instance, PCR reaction can be used to test for the presence of particular STS's in slices.

Genomic DNA from a monosomic hybrid cell line containing human chromosome 20 was used in these experiments. STSs previously located and mapped onto chromosome 20 were used to order the cloneless library fractions (DNA in gel slices). These experiments mapped the *Not* I restriction fragments on chromsoomes 20 with at least an order or magnitute increase in efficiency than in similar efforts. Furthermore,

unlike conventional mapping experiments using hybridization, the results linked each STS to a source of genomic DNA that could use in additional experiments. For instance, the cloneless library fractions from the region amplified in cancer was used as a hybridizaton probe to as probe to isolate hncDNAs. Long inter-*Alu* PCR was used to amplify and ^{32}P-label human DNA from the cloneless libary fractions. The labeled DNA was used as a hybridization probe to an arrayed hncDNA library. About ninety clones were identified and sequenced that hybridized to this region. Other available genomic resources (e. g. cloned sequences) were also used as hybridization probes. About 10 STS PCR primers were designed, and gel slices and available large insert clones in the amplified region were tested for the occurrence of the selected sequences. The results of these experiments indicate that eight of the ten sequences come from the selected chromosomal region. This confirms other experiments done in collaboration with Joe Gray using FISH (flourescent *in situ* hybridization) that demonstrated that the cloneless library fractions provided regional specific DNA. Most recently we have explored the best way of amplifying the genomic DNA in the slices so that the template DNA supply from a single experiment can be used in many applications. A long term goal of these experiments is use the cloneless library fraction to make up arrays that can be used in experiments similar to, but easier than, CGH.

Summary

Great technical progress has been made with our developing methods. Several articles are being written up now which focus mostly on the methodology. However, we have now begun to apply the methods to a small number of breast cancer tumor cells to identify the problems that are posed by the pecularities of those samples

The major progress on genomic profiling entails the realization that the originally proposed method is not as powerful as newly developing MS methods. Thus, we decided to take a very forward looking approach to cDNA profiling. Fortunately, the basic methods of DNA handling are almost the same as those proposed in the original grant. Specific adaption to MS is now being done. Meanwhile several other methods for speeding gene searches have also been developed.

References

Brenner, S. and Livak, K. J. 1989. Proc. Nat. Acad. Sci., 86: 8902-8906.

Broude, N., Sano, T., Smith, C. L., and Cantor, C. R. 1994. Proc. Nat. Acad. Sci. USA 91: 3072-3076.

Broude, N., Chandra, A., and C. L. Smith. 1997. Proc. Nat. Acad Sci. USA 94: 4548-4553.

Broude, N., Lykyanov, S. A., Graber, J. H., Sverdlov, E. D., and Smith, C.L. 1997. Submitted.

Ermolaeva, O. D. and Sverdlov, E. D. 1995. Genetic Analysis (Biomolecular Engineering) 13: 49-54.

Fu, D. J., Broude, N. E., Koester, H, Smith, C. L., and Cantor, C. R. 1995. Genetic Analysis (Biomolecular Engineering) 12: 137-142.

Harshman, K., Bell, R., Rosenthal., J., *et al.*, 1995. Hum. Mol. Genet. 4: 1259-1266.

Kallioniemi, A., Kallioniemi, O.-P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W., and Waldman, F. M. 1994. Proc. Natl. Acad. Sci. USA 91: 2156-2160.

Kato, K. 1995. Nucleic Acids Res. 23: 3685-3690.

Kato, K. 1996. Nucleic Acids Res. 24: 394-395

Liang, P. and A. B. Pardee 1992. Science 257: 267

Liang, P., Zhu, W., Zhang, X., Guo, Z., O'Connell, Averbukh, L., Wang, F. and Pardee, A. 1994. Nucleic Acids REs. 22, 5763-5764.

Lisitzyn, N., Lisityn, N., and Wigler M. 1993a. Science 259, 946-951.

Lisitzyn, N., Rosenberg, M., Launer, G., Wagner, L., Potapov, V., Kolesnik, T.,

and Sverdlov, E. 1993b. Mol. Genet. Microbiol. and Virusol. 3, 26-29.

Sager, R, Sheng, S., Anisowixz, A., Sotiropouloou, G., Zou, A., Stenman, G., Swisshelm, Z., Chen, M., Hendrix, J.C., Pemberton, P., Rafidi, K., and Ryan, K. 1994. Cold Spring Harbor Symposia on Quant. Bio. LIX: 537-545.

Schena, M., Shalon, D., Davis, R. W., and Brown, P.O. 1995. Science 270, 467-70

Smith, D. R. 1992. PCR Meth. Appl., 2: 21-27

Tanner, M., M. Tikkonen, A Kallioniemi, C. Collins, T. Stokke, R. Karhu, D. Kowbel, F. Shadravan, M. HIntz, W.L. Kuo, f. M. Waldman, J. Isola and J. W. Gray. 1994. Cancer Research 54: 4257-4260.

Tavtigian, S. V., Simard, J., Rommens, J., *et al.* 1996. Nat. Genet. 12: 333-337.

Unrau, P. and Deugau, K. V. 1994. Gene 145: 163-169.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. 1995. Science 270: 484-487.

Publications

Bouchard, J., M. Lewis, J. Graber, N. Storm, C. L. Smith. 1998. Enhancing Targeted and cDNA Differential Display using a *Saccharomyces cerevisiae* model system, in preparation.

Broude, N., A. Chandra, and C. L.Smith. 1997. Differential Display of Genome Subsets Containing Specific Interspersed Repeats. Proc. Nat. Acad. Sci. USA 94: 4548-4553.

Broude, N., N. Storm, S. Malpel, J. H. Graber, S. Lykyanov. E. Sverdlov and C. L. Smith. 1998. PCR Based Targeted Genomic and cDNA Differential Display. Biomolecular Engineering, in press.

Christine Debouck. 1995. "Differential display or differential dismay?" Current Opinion in Biotechnology, 6:597-599

Collins, C., J. M. Rommens, D. Kowbel, T. Godfrey, M. Tanner, S.-I. Hwang, D. Polikoff, G. Nonet, J. Cochran, K. Myambo, K. E. Jay, J. Froula, T. Cloutier, W.-L. Kuo, P. Yaswen, S. Dairkee, J. Giovanola, G. B. Hutchinson, J. Isola, O.-P Kallioniemi, M. Palazzolo, C. Martin, C. Ericsson, D. Pinkel, D. Albertosn, W.-B. Li, J. W. Gray. 1998. Positional Cloning of ZNF217 and NAC1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. Proc. Nat. Acad. Sci. USA, Vol. 95, pp. 8703-8707

Foulon, C., Z. Nyugen, N. Storm and C. L. Smith. 1998. Genomic Discordance in Mongozygotic Twins. In, "Techniques in the Behavioral and Neural Sciences", Eds. W. Crusio and R. Gerlai, Elsevier, Amsterdam, The Netherlands.

Goffean, A., B. Barrell, H. Bussey, R. Davis, B. Dujon, H. Feldmann, F. Galiber, J. Hoheisel, C. Jacq, M. Johnsto, E. Louis, H. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. Oliver. 1996. Life with 6000 genes. Science 274, 546-567

Graber, J. H, D. J. Fu, H. Koester, C. L. Smith, and C. R. Cantor. 1997. Optimizing Spectrometry Based Genetic Mutation Localization and Identification. Mathematical Modelling and Scientific Computing. in press

Graber, J. H., O'Donnell, M. J., Smith, C. L. and Cantor, C. R. 1998. Advances in DNA diagnostics. Current Biology. in press
Graber, J., C. R. Cantor C. L. Smith 1998. Genetic Mutation Screening by Mass Spectrometry. Genetic Analysis (Biomolecular Engineering). in press

Hatch, A., T. Sano, J. Misasi and C. L. Smith. 1998. Mutation Detection using Rolling Circle Amplification (RCA) and Immobilized DNA probes. Genetic Analysis (Biomolecular Engineering) in press.

Lavrentieva, I., N. Broude, Y. Lebedev, I. I. Gottesman, S. A. Lukyanov, E. Sverdlov, and C. L. Smith. 1998. Genomic DNA Fingerprinting of Human Endogenous Retrovirus LTR in Monozygotic Twins. submitted.

Maas, D., M. Lewis, N. Bukanov and C. L Smith. 1998. Cloneless Libraries: Long Range Mapping of Chromosome 20 and cDNA Selection for The 20q13 Amplified in Cancer Cells, in preparation

Oliveira, R. P., N. E. Broude, A. M. Macedo, C. R. Cantor, C. L. Smith and S. D. J. Pena. 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Nat. Acad. Sci. 95: 3776-3780.

Storm, N., S. Malpel, M. Lewis, Y. Shevchenko, N. Broude, and C. L. Smith. 1998. Optimizing Targeted Genomic and cDNA Genomic Differential Display, in preparation

Surdi, G., R. Yaar, and C. L. Smith. 1998. Discrimination of DNA Duplexes with Matched and Mismatched Tandem Repeats by T4 Endonuclease. Genetic Analysis (Biomolecular Engineering), in press.

Yaar, R., P. Szafranski, C. R. Cantor and C. L. Smith. 1996. *In situ* Detection of Tandem Repeat Length" Genetic Analysis (Biomolecular Engineering) 13: 113-118.

Researchers Supported

Cassandra Smith	9/94 - 9/98
Nikolay Bukanov received	1/2 paycheck on 1/31/96
Natalia Broude	5/31/96 - 10/31/96 100%
	5/31/97 - 01/31/98 50%
	1/31/98 - 04/23/98 100%
Sarah Malpel	1/31/97 - 12/31/97 100%
Yuri Shevchenko	1/31/97 - 04/30/97 100%
David Maas	12/29/95 - 10/4/96 100%
Walid Soussou	12/29/95 - 5/23/97 100%
John Misasi	08/02/96 - 09/5/97 100%

Anson Hatch 11/27/97 - 7/24/98 100%